

Oracle in der Cloud Hochdruckgebiet oder Regenwolken?

Andreas Stephan
Bayer Business Services GmbH
Leverkusen/Deutschland

Schlüsselworte: Cloud Computing, Oracle, Database, Amazon, RDS, AWS, AMI

Einleitung

Der Cloud-Hype ist in der realen Welt angekommen. Ernstzunehmende Lösungen werden von den Anbietern auf dem Markt positioniert. Datenbanken wie Oracle und MySQL können auf Stundenbasis gemietet werden und versprechen bei extrem kurzer Deployment-Zeit eine hochflexible Umgebung zu geringsten Kosten. Kann dieses Versprechen gehalten werden? Welche Risiken gehen wir ein, wenn wir unsere Daten ins Internet stellen? Wie kann man Datenbanken wie Oracle 11g in der Cloud betreiben, welche Herausforderungen stellen sich uns und (wie) kann man diese bewältigen? In diesem Vortrag wird versucht, diese Fragen zu beantworten, was je nach Anforderung an Kosten, Verfügbarkeit, Vertraulichkeit und Sicherheit der Daten unterschiedliche Ergebnisse bringen kann.

Warum eigentlich Cloud Computing?

Es gibt mehrere Gründe, sich mit Cloud Computing zu beschäftigen. Die eher trivialen Gründe, wie „macht Spaß“ und „interessantes Thema“ lassen wir einmal beiseite und widmen uns den eher naheliegenden Gründen „Kostensparnis“ und „Geschwindigkeit“.

Alle Anbieter von IT-Services – auch die internen Serviceprovider größerer Unternehmen – sind naturgemäß durch ihre Kunden einem ständigen Kostendruck unterworfen und gezwungen, ihre Services zu immer geringeren Kosten (und Preisen) anzubieten. Jede Möglichkeit, die Kosten bei gleichbleibender Qualität zu reduzieren, erscheint daher ein probates Mittel.

Kann man mit Cloud Computing Geld sparen? Eine Frage, die man mit Radio Eriwan beantworten kann: Im Prinzip ja, aber... Dieses „aber“ sollte man jedoch gut im Auge behalten, um ein böses Erwachen zu vermeiden. Dazu jedoch später mehr im Detail.

Unmittelbar verbunden mit der Kostensparnis ist die Geschwindigkeit. Hatte man früher einige Wochen Zeit, für ein neues IT-Projekt die Hardware für Datenbank und Applikations-Server zu beschaffen, zu installieren, zu testen und dem Kunden bereitzustellen, reduzieren sich heutzutage die Bereitstellungszeiträume auf Tage oder idealerweise auf Stunden. Ist man in der glücklichen Lage jederzeit über genügend freie Hardware-Ressourcen zu verfügen, sollte das keine allzu großen Probleme aufwerfen, sofern man seine Hausaufgaben gemacht hat und funktionierende (halb-) automatische Deploymentmechanismen produktiv hat. Nur leider sind in Zeiten von immer enger werdenden IT-Budgets freie Infrastruktur-Ressourcen totes Kapital, was weder die Shareholder noch das Management erfreuen dürfte.

Der Ausweg aus dieser Zwickmühle bietet sich dadurch, dass man den Ressourcen-Pool eines darauf spezialisierten Anbieters nutzt („Infrastructure as a Service“ - IaaS) und somit Ressourcen schnell und genau dann, wenn man sie benötigt, nutzen kann. Wenn es dann noch gelingt, die Standardisierung der (Datenbank-) Infrastruktur soweit vorwärts zu treiben, dass man fertige Images nahezu unverändert nutzen kann, ist man dem Ziel eines Deployments „auf Zuruf“ schon sehr nahe.

Kostenreduktion

Damit man feststellen kann, ob Cloud Computing Kosten sparen kann, muss man zuerst einmal den eigenen Kostenapparat einer Analyse unterziehen und die Kosten so aufschlüsseln, dass sie vergleichbar werden. Einge kaufte Cloud-Services von z.B. Amazon haben im Allgemeinen einen deutlich anderen Leistungsumfang, als z.B. der eigene, „traditionelle“ Datenbankservice. Letzterer ist eine über Jahre entwickelte komplexe Dienstleistung mit Hard- und Software, Lizenzen, Backup, Monitoring, Upgrades, Patches, Virens scanner, Firewall, Security-Mechanismen usw. Der Cloud-Service umfasst im Allgemeinen nur die reine Bereitstellung von Hard- und Software, eventuell wie z.B. bei Amazons RDS, auch die Datenbank. Dabei kann man noch unterscheiden, ob die Lizenz enthalten sein soll oder über ein „Bring your own licence“ Modell selber bereitgestellt werden muss. Eingebettete Lizenzen sind allerdings meist beschränkt auf spezielle Editionen der Datenbank – im Beispiel von Amazons RDS Service ist es die Standard-Edition One, die gegenüber der Standard- und der Enterprise Edition deutliche Einschränkungen aufweist.

Der Vorteil von Cloud-Services verschwindet recht schnell, wenn die obigen additiven Leistungen (Monitoring, Help-Desk, Backup/Recovery, Security etc.) addiert werden. Trotzdem kann sich ein Cloud-Service durchaus rechnen. Es kommt dabei auf mehrere Faktoren an. Zum einen muss man bei Cloud-Services die Übertragungskosten ins und aus dem Netz des Serviceproviders zum Servicepreis addieren, was i.A. im Voraus nicht einfach fällt, da meistens entsprechende Statistiken für eine Anwendung nicht vorliegen. Zum anderen werden Cloud-Services üblicherweise neben Fixkosten für die notwendige Hardware (Memory, CPUs, Storage) auch mit einem zeitlich basierten Faktor abgerechnet. Jede Stunde Laufzeit einer Cloud-Datenbank kostet dann einen entsprechenden Betrag. Diese variablen (Daten-) und zeitabhängigen Kosten werden üblicherweise in einem internen IT-Service Unternehmen so nicht abgerechnet, sondern sind anteilig in den Fixkosten eines Systems (z.B. Housingkosten, Kosten für Netzwerkports, etc.) enthalten. Stellt man einmal die Kosten für eine „Full-Use“ Datenbank, also 7x24h oder 5x9h zusammen und kalkuliert bei dem Cloud-Anbieter seiner Wahl die Kosten für Netzwerktransfer hinzu, sowie die Kosten für die additiven Services, wie Monitoring, Backup, Helpdesk etc. stellt man fest, dass für solche Anwendungen eine Cloud-Lösung schnell teurer wird, als sie intern zu betreiben. Für Kleinbetriebe oder Mittelständler kann sich das trotzdem durchaus rechnen, da die interne Serverlandschaft nicht groß genug ist, so dass die anteiligen Kosten für Klima, Backup, (Not-)Strom, Housing etc. die Kalkulation zu Gunsten der Cloud-Services ausgehen lässt.

Ganz anders sieht die Kalkulation aus, wenn man die Datenbank nur für eine begrenzte Zeit benötigt, oder nur für eine bestimmte Zeit mit hoher Last betreibt und den Rest der Zeit mit eingeschränkten Ressourcen auskommt. Als Beispiel seien hier Abrechnungssysteme genannt, die nur einmal im Monat oder Quartal hohe Anforderungen an CPU und Memory haben weil Datenverdichtung und/oder Reporting laufen. Für diesen Zweck kann eine Cloud-DB per Mausclick um CPUs und Memory erweitert werden, um nach getaner Arbeit wieder auf das notwendige Minimum reduziert zu werden. Bei selbst betriebener, eigener Hardware fällt das schwerer, auch wenn es nicht unmöglich ist. Jedoch birgt die naturgemäß begrenzte Ausstattung an Hardware immer die Gefahr, die Erweiterung nicht durchführen zu können, da gerade keine freien Ressourcen zur Verfügung stehen. Hier schlägt das Pendel also ganz eindeutig in Richtung Cloud-Services aus.

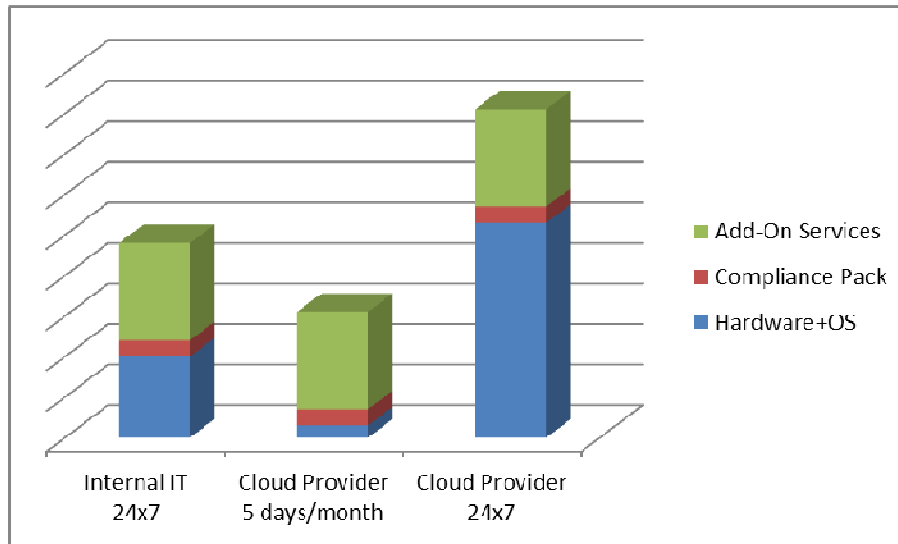


Abbildung 1: Kostenvergleich interne vs. Cloud Services

Wie man in der obigen Grafik erkennen kann, ist die reine Bereitstellung von Infrastruktur durch einen Cloud-Service Provider erheblich günstiger, als durch die interne IT. Durch die Zeitabhängigkeit des Cloud-Services gilt das jedoch nur bis zu einer bestimmten Anzahl an Betriebsstunden. Danach ist die interne IT wieder günstiger, da hier keine dynamischen Komponenten (Laufzeit, Traffic) existieren.

Geschwindigkeit

Ein unbestreitbarer Vorteil der Cloud Services ist die Geschwindigkeit der Bereitstellung. Da die Cloud-Anbieter voll automatisierte Self-Service-Anwendungen bieten, muss man nicht auf den DBA oder SYSADMIN warten, der einem das System aufsetzt und bereitstellt. Im Prinzip könnte man als Anwender die Datenbank selber mittels weniger Klicks auf Amazons AWS- oder RDS-Seiten bereitstellen. In der Praxis gestaltet sich das jedoch etwas schwieriger, da hier doch ein gutes Maß Wissen über Security, Firewalls, Netzwerke, CPUs und Speichergrößen notwendig ist.

Durch die Verlagerung bzw. Voraussetzung dieses Wissen und der notwendigen nachfolgenden Tätigkeiten zum Kunden hin, kommt man dem Ziel der „On-Demand“ Lösung schon sehr nahe – zumindest erscheint es dem geneigten Kunden so. Tatsächlich muss er nun die kompletten Arbeitsabläufe selber durchführen, die sonst sein Serviceprovider für ihn erledigt hat. Logging der Aktivitäten, Change-Management, Dokumentation, einbinden in Backup und Monitoring Umgebungen, Netzwerkkomponenten konfigurieren, Security Scan sowie Test und QA sind alles Aktionen die einen Großteil der Arbeit eines DBAs beim Deployment einer neuen Datenbank ausmacht. Das Weglassen dieser Aktivitäten bzw. Verlagerung zum Kunden bedeutet natürlich zum einen eine erhebliche Vereinfachung des Deployments und damit eine hohe Fähigkeit zur Automatisierung, aber eben auch einen immensen Preisvorteil, da die teilweise manuellen Tätigkeiten vergleichsweise teuer sind. Möchte man aber eine solche Cloud-Datenbank unter den gleichen Gesichtspunkten betreiben, wie vorher eine Datenbank durch die interne IT, also gleiche Anforderungen an Verfügbarkeit, Wiederherstellbarkeit, Sicherheit, etc. so stehen nach dem extrem schnellen Deployment der DB die anderen, manuellen Tätigkeiten ebenfalls an – und kosten Geld und Zeit, was den Geschwindigkeitsvorteil zumindest teilweise zunichtemacht.

Datenschutz und -sicherheit

Eine der ersten Fragen, die bei einer Diskussion über Cloud-Anwendungen gestellt wird, ist etwas in der Art „Wo sind denn eigentlich meine Daten“ oder „Sind meine Daten sicher?“. Die Antwort auf diese Fragen fällt nicht leicht, da Cloud-Anbieter üblicherweise mehrere auf dem Erdball verteilte Rechenzentren betreiben, in denen die Cloud-Datenbanken liegen könnten. Neben den deutschen Datenschutzgesetzen, die den Umgang mit personenbezogenen Daten regeln, existieren ja auch noch die Notwendigkeiten der Ersteller von Daten, diese geheimzuhalten, bis z.B. die klinische Erprobung eines Medikamentes abgeschlossen, die Patente eingereicht, bzw. das Produkt auf dem Markt ist. Hier ist man zuerst einmal gefordert, eine Klassifizierung der Daten vorzunehmen – z.B. in „Public“, „Confidential“ und „Personal“ – und abhängig von der Klasse und den geltenden Rechten (Datenschutzbedingungen) die Ablage der Daten z.B. in einer öffentlich erreichbaren Cloud DB zu erlauben, diese in einer Virtual Private Cloud unterzubringen oder lokal im eigenen Rechenzentrum zu betreiben. Prinzipiell schadet es nicht, wenn man den worst-case annimmt und seine Daten stark verschlüsselt, bevor man sie auf einem Cloud Server ablegt. Hierzu bietet Oracle ja einige Funktionen an, die es erlauben, sowohl die Datenbankdateien auf den Cloud-Servern, als auch den Datentransfer komplett zu verschlüsseln. Bei öffentlich erreichbaren und somit potentiell einer höheren Gefährdung unterliegenden Cloud-Services sollte man ebenfalls darauf achten, die regelmäßig erscheinenden Security Patches einzuspielen, um einer Ausnutzung der Sicherheitslücken vorzubeugen. Natürlich bestehen diese Gefahren auch im Intranet eines Unternehmens, nur da hat man durch den begrenzt möglichen Zugang von außen, eine bessere Überwachung von Einbruchversuchen.

Hinzu kommen noch die Bestrebungen verschiedener Regierungen, die ein bestimmtes Interesse an sämtlichen Daten haben, sei es personenbezogene oder produktspezifische. Hier gibt es dann z.B. in den USA die Gesetzeslage (US Patriot Act), die amerikanische Cloud Services Anbieter in Europa, wie z.B. Microsoft, Amazon, Google oder Oracle) dazu zwingen, auch ohne Nachfrage den jeweiligen Regierungs-Institutionen/Geheimdiensten zur Verfügung zu stellen. Dies widerspricht zwar den Datenschutzregeln und –Rechten der EU, hindert aber die USA nicht daran, so zu verfahren. Hat man somit bisher geglaubt, die wichtigen Daten in den Cloud-Rechenzentren in Europa sicher untergebracht zu haben, so sollte man sicherheitshalber einmal die Datenschutzbestimmungen der Provider, bzw. deren Ausnahmen genau durchlesen.

Eine weitere Schutzstufe ist die Virtual Private Cloud, die eine gewisse Abschottung der Datenbankinhalte gegen das normale Internet bietet. Hier dient die Cloud bzw. deren Services als eine Art erweiterte Blase des eigenen Rechenzentrums. Die Dienste, z.B. Datenbanken sind nur von innerhalb des eigenen Netzes erreichbar und werden vom Anbieter mit hohen Schutzmechanismen gegen Zugriffe vom Internet her geschützt. Zudem können dann die Infrastrukturkomponenten wie z.B. Server in den gleichen IP-Adressräumen wie das interne Netz genutzt werden. Das heißt auch, dass interne Leistungsmerkmale wie Backup, Monitoring, Security Scan etc. durch die Einbindung ins eigene Netz genutzt werden können, was bei der Public Cloud nicht ganz so einfach ist, da man im Allgemeinen den Zugriff vom Internet ins eigene Intranet möglichst weitgehend unterbindet, um nicht die Anzahl der Angriffspunkte heraufzusetzen.

In letzter Zeit sind Cloud-Provider bzw. deren Kunden wiederholt von Ausfällen und sogar Datenverlusten betroffen gewesen. Hier muss man als Kunde zur Zeit offenbar noch gewisse Abstriche in Bezug auf die Verfügbarkeit der Daten machen und die notwendigen Verfahren und Methoden für ein geregeltes Backup und Recovery einführen, so wie man es bei Inhouse Anwendungen auch macht. Die Cloud-Provider bieten einige unterschiedliche Methoden für die Erhöhung der Datenverfügbarkeit an, die man auch tunlichst nutzen sollte, sofern man seine Daten

nicht in angemessener Zeit von den Quell-Systemen wieder herstellen kann. Backups, Snapshots und die Verteilung bzw. Spiegelung der Anwendung und deren Daten auf verschiedene Availability Zonen sind z.B. bei Amazon eine Möglichkeit drohendem Datenverlust vorzubeugen.

Fazit

Die Nutzung von Cloud-Services kann durchaus erhebliche Vorteile bringen. Geschwindigkeit, dynamische Nutzung von CPU und/oder Memory, variable Kosten auf Basis der Nutzungsdauer und -intensität, sowie der nahezu unbegrenzte Vorrat an (virtueller) Infrastruktur versprechen hochflexiblen Betrieb bei geringsten Kosten. Rechnet man jedoch die notwendigen Zusatzleistungen für einen normalen IT-Infrastrukturbetrieb hinzu, ohne den man vernünftigerweise keine Datenbank betreiben würde, rechnet sich zur Zeit der Datenbankbetrieb in der Cloud nur für Anwendungen, die keine „Standard“ Betriebszeiten 5x9 oder 7x24 benötigen. Vor dem Hintergrund weiter fallender Preise könnte aber auch diese Einschränkung bald fallen. Die hochflexible Handhabung von IT-Ressourcen, wie z.B. die extrem schnelle Bereitstellungszeit für einen DB-Server, erlaubt aber auch andere Anwendungsmöglichkeiten, von denen DBAs bisher nur träumen konnten. Für ein Recovery innerhalb kürzester Zeit und zu minimalen Kosten eine eigene Datenbank samt OS, Memory und Storage aufzubauen und ohne Beeinträchtigung des laufenden Produktionsbetriebes der Original-DB Daten zu restaurieren, zu testen und wieder bereitzustellen war bisher nur schwer möglich.

Die Wichtigkeit und Schutzwürdigkeit der Daten sollte man bewerten und auf Basis dieser Bewertung Regeln entwerfen, nach denen Daten(banken) in die Cloud dürfen oder eben nicht. Vor diesem Hintergrund sind datenschutzrechtliche Probleme, wie die Weitergabe der Daten an Geheimdienste ohne Information des Daten-Eigentümers und gegen geltendes Recht eine weitere Stufe im eigenen Bewertungsplan.

Die Großwetterlage für Cloud-Datenbankservices ist also heiter bis wolkig, mit einigen möglichen Regenschauern. Aber auch hier gilt: wer sich passend anzieht, kommt auch trocken durch einen Regenguss.

Kontaktadresse:

Andreas Stephan

Bayer Business Services GmbH
C152/1
D-51368 Leverkusen

Telefon: +49 (0) 214-30 31411
E-Mail andreas.stephan@bayer.com
Internet: www.bayerbbs.com